



EEG Sequence Imaging: A Markov Prior for the Variational Garrote

Hansen, Sofie Therese; Hansen, Lars Kai

Published in:

Proceedings of the 3rd NIPS Workshop on Machine Learning and Interpretation in Neuroimaging 2013

Publication date:

2013

[Link back to DTU Orbit](#)

Citation (APA):

Hansen, S. T., & Hansen, L. K. (2013). EEG Sequence Imaging: A Markov Prior for the Variational Garrote. In *Proceedings of the 3rd NIPS Workshop on Machine Learning and Interpretation in Neuroimaging 2013*

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

EEG Sequence Imaging: A Markov Prior for the Variational Garrote

Sofie Therese Hansen and Lars Kai Hansen

Section for Cognitive Systems, DTU Compute, Technical University of Denmark,
Kgs. Lyngby, Denmark
sofha@dtu.dk and lkai@dtu.dk, home page: www.compute.dtu.dk

Abstract. We propose the following generalization of the Variational Garrote for sequential EEG imaging: A Markov prior to promote sparse, but temporally smooth source dynamics. We derive a set of modified Variational Garrote updates and analyze the role of the prior’s hyperparameters. An experimental evaluation is given in simulated data and in a benchmark EEG data set.

Keywords: Source reconstruction, Bayesian inference, the Variational Garrote, EEG, sparse Bayesian learning.

1 Sparse Sequence Reconstruction

The dynamics of electroencephalographic (EEG) sources is an active research field, see. e.g., [1–6]. We are interested in the spatio-temporal source distribution under well-defined brain activation. The contribution of the present paper is to expand upon a promising new algorithm, the so-called Variational Garrote (VG), first proposed in [7] and recently applied to EEG brain imaging in [8] and expanded to a fixed sparsity temporal model in [9]. Our goal and main contribution in this presentation is to relax the fixed sparsity model by introducing a flexible Markov prior forming a new algorithm that we refer to as MarkoVG.

The forward relation between cortical sources and electrode potentials is linear and we will here assume the forward propagation model known, although attempts have been made of estimating it from data, see e.g. [5]. Using a ‘spike and slab’ like representation, the linear relation between observations across time, $\mathbf{Y} \in \mathbb{R}^{K \times T}$, the forward model $\mathbf{A} \in \mathbb{R}^{K \times N}$ and the source matrix $\mathbf{X} \in \mathbb{R}^{N \times T}$ is modified in the Variational Garrote (VG) [7] as

$$Y_{k,t} = \sum_{n=1}^N A_{k,n} X_{n,t} + E_{k,t} \xrightarrow{\text{VG}} Y_{k,t} = \sum_{n=1}^N A_{k,n} S_{n,t} X_{n,t} + E_{k,t} , \quad (1)$$

where $S_{n,t}$ is a 0,1 binary variable controlling the spatio-temporal support of brain activity $X_{n,t}$ (i.e., at the dipolar location n at time t). The variable $E_{k,t}$ is assumed to be i.i.d. normal noise with zero mean and unknown variance $1/\beta$. Aiming for temporally smooth and spatially sparse configurations we assign a

simple Markov model prior for the binary variables of a specific dipole location n , represented by a transition matrix $\Gamma_{i,j} = \text{Prob}(S_{n,t} = i | S_{n,t-1} = j)$ with two free parameters, e.g., of the form

$$\Gamma = \begin{bmatrix} \Gamma_{0,0} & \Gamma_{0,1} \\ \Gamma_{1,0} & \Gamma_{1,1} \end{bmatrix} = \begin{bmatrix} 1 - \Gamma_{1,0} & \Gamma_{0,1} \\ \Gamma_{1,0} & 1 - \Gamma_{0,1} \end{bmatrix} . \quad (2)$$

The stationary distribution of Γ is given by $(\Gamma_{0,1}/(\Gamma_{1,0} + \Gamma_{0,1}), \Gamma_{1,0}/(\Gamma_{1,0} + \Gamma_{0,1}))$, thus the ratio $\frac{\Gamma_{0,1}}{\Gamma_{1,0}}$ controls the prior sparsity.

The VG approach is based on approximate variational inference. Here we derive the modified update rules for the variational approximation. With uniform priors on \mathbf{X} and β we obtain a variational free energy $F(q, \mathbf{X}, \beta)$ which is minimized to obtain the optimal variational distribution q , the source estimates \mathbf{X} , and the noise parameter β . To reduce computation we use the *dual formulation* [7] introducing $Z_{k,t} = \sum_n A_{k,n} M_{n,t} X_{n,t}$ and Lagrange multipliers $\lambda_{k,t}$

$$F = -\frac{KT}{2} \log \frac{\beta}{2\pi} + \frac{\beta}{2} \sum_{t,k} (Y_{k,t} - Z_{k,t})^2 + \frac{K\beta}{2} \sum_{t,n} M_{n,t} (1 - M_{n,t}) X_{n,t}^2 \chi_{n,n} \quad (3)$$

$$- \sum_{n,t} \left[M_{n,t} \log \frac{\Gamma_{1,0}}{\Gamma_{0,0}} + M_{n,t-1} \log \frac{\Gamma_{0,1}}{\Gamma_{0,0}} + (M_{n,t} M_{n,t-1}) \log \frac{\Gamma_{0,0} \Gamma_{1,1}}{\Gamma_{0,1} \Gamma_{1,0}} \right] \quad (4)$$

$$+ NT \log \frac{1}{\Gamma_{0,0}} + \sum_{n,t} [M_{n,t} \log(M_{n,t}) + (1 - M_{n,t}) \log(1 - M_{n,t})] \quad (5)$$

$$+ \sum_{t,k} \lambda_{k,t} \left(Z_{k,t} - \sum_n A_{k,n} M_{n,t} X_{n,t} \right) , \quad (6)$$

where $\chi = \mathbf{A}^T \mathbf{A} / K$. The variational estimates satisfy the following equation set (with $\sigma(x) = (1 + \exp(-x))^{-1}$)

$$X_{n,t} = \frac{1}{K\beta} \frac{1}{(1 - M_{n,t}) \chi_{n,n}} \sum_k \lambda_{k,t} A_{k,n} , \quad Z_{k,t} = Y_{k,t} - \frac{1}{\beta} \lambda_{k,t} , \quad (7)$$

$$\beta = \frac{1}{TK} \sum_{t,k,c} \lambda_{k,t} \lambda_{c,t} C_{k,c,t} , \quad C_{k,c,t} \equiv \delta_{k,c} + \frac{1}{K} \sum_n \frac{M_{n,t}}{(1 - M_{n,t}) \chi_{n,n}} A_{k,n} A_{c,n} \quad (8)$$

$$\lambda_{c,t} = \beta \hat{Y}_{k,t} \text{ with } \sum_c C_{k,c,t} \hat{Y}_{k,t} = Y_{k,t} , \quad (9)$$

$$M_{n,t} = \sigma \left(\frac{K\beta}{2} \chi_{n,n} X_{n,t}^2 + \gamma_1 + \gamma_2 (M_{n,t-1} + M_{n,t+1}) \right) , \quad (10)$$

solved by iteration. Here the combination of Markov parameters; $\gamma_1 = \log \left(\frac{\Gamma_{1,0} \Gamma_{0,1}}{\Gamma_{0,0}^2} \right)$, $\gamma_2 = \log \left(\frac{\Gamma_{0,0} \Gamma_{1,1}}{\Gamma_{0,1} \Gamma_{1,0}} \right)$ determine the sparsity and smoothness of the solution: The parameter γ_2 is seen to control the degree of temporal smoothness, while γ_1 corresponds to Kappen's sparsity control parameter (with negative values favoring sparse solutions). If $\gamma_2 = 0$ we recover the original VG algorithm.

2 Experimental Evaluation

In the following a simulation example will serve to illustrate the properties of MarkoVG. We simulate a data set of size $N = 150$, $K = 25$ and $T = 25$. The weight distribution is controlled as three active sources with sine wave like time courses active at different time windows. The data are corrupted by normal additive noise (SNR= 5 dB). An example of the generated sources can be seen in the left panel of Fig. 1, while the right panel shows the estimated sources using MarkoVG. Here the parameter $\Gamma_{0,1} = 0.02$ is fixed, while the ratio $\frac{\Gamma_{0,1}}{\Gamma_{1,0}}$ is estimated through three-fold cross-validation among possible values ranging from 10^{-5} to 10, in 50 steps. For each step 25 iterations are performed.

We find that the relevant weights are recovered while one irrelevant weight (at time $t = 11$) is mistakenly judged as being relevant. Swift convergence of $M_{n,t}$ is seen in Fig. 2 for both the 'true' locations (depicted in blue, green and red corresponding to Fig. 1) and for one 'false' location, chosen as the location with non-zero activity at time $t = 11$ (depicted in gray). The color indicates the value of $M_{n,t}$; brightest or darkest indicate $M_{n,t}=0$ or 1 (minimal or maximal marginal posterior probability of activation in location n). In Fig. 2, MarkoVG is further evaluated using the source retrieval score (F_1 -measure) and the mean squared error (MSE) on the estimated sources. A total of 100 randomly generated data sets similar to that shown in Fig. 1 are drawn. The performance measures are seen to converge quickly towards their optimum. Note that the iterations minimize the Free energy whose optimum is not simply related to source retrieval (F_1) nor MSE. Varying both γ_1 and γ_2 in a grid to inspect the influence of the smoothness and sparsity parameters on the F_1 -measure and MSE we obtain Fig. 3. Note 'hot' colors indicate high values (better performance in F_1 -measure and worse for the MSE). A (γ_1, γ_2) -region exists with high source retrieval ability and low MSE.

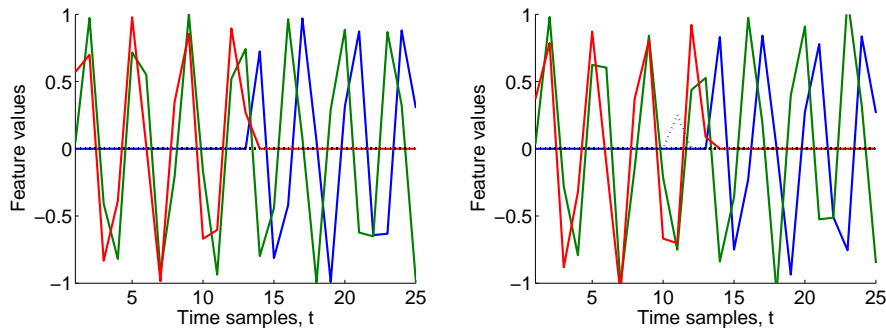


Fig. 1: The source time functions, true (left) and estimated (right). The active sources are in blue ($n=1$), green ($n=2$) and red ($n=3$) respectively, and the non-active sources ($n=4:150$) are all represented as black dotted lines. Three-fold cross-validation is used to find an optimum level of $\Gamma_{0,1}/\Gamma_{1,0}$. For each investigated level 25 iterations are applied. SNR= 5 dB.

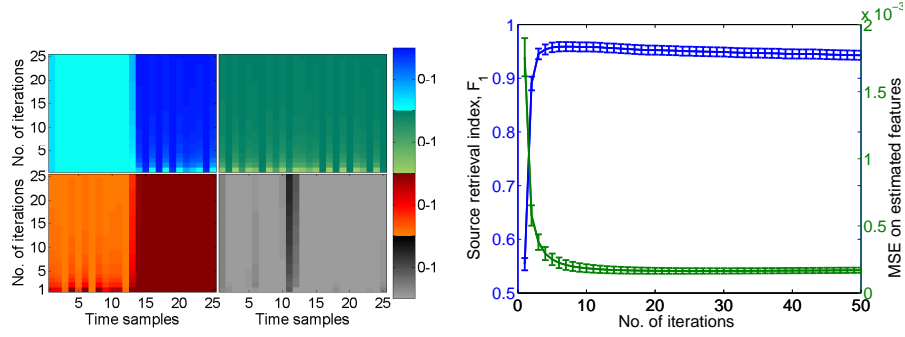


Fig. 2: Left: Convergence of $M_{n,t}$ during 25 iterations for the 3 planted active sources (blue, green and red) and the strongest false source (gray scale). Example corresponds to that shown in Fig. 1. Right: Evaluation of MarkoVG with 100 repetitions of data sets similar to that shown in Fig. 1, all with SNR around 5 dB. Performance shown as mean \pm standard deviation of source retrieval F_1 -measure and MSE on the weights.

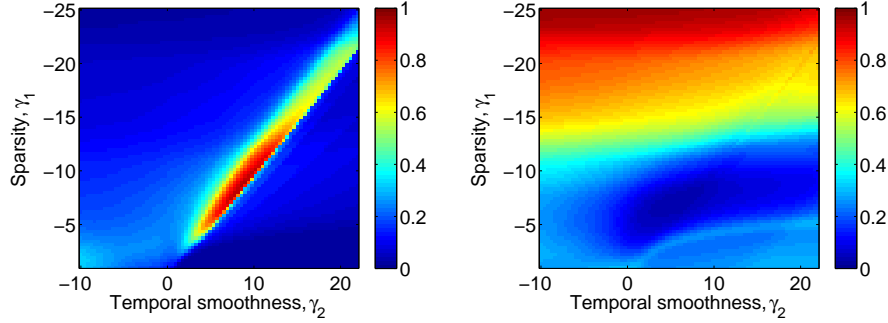


Fig. 3: Search for optimal smoothness and sparsity. Left: mean F_1 , right: mean MSE, across 30 repetitions of data sets similar to Fig. 1(left). For each parameter combination 25 iterations are applied. SNR= 5 dB.

The existence of this region indicates that the MarkovVG representation indeed allows us to find sparse sources with limited bias on the source magnitudes.

The performance of MarkoVG is next examined on benchmark EEG data which is part of a multi-modal face response data set, available through the SPM website¹. The data used here are collected from a single subject and used to demonstrate the modulation in brain activity when seeing faces vs. scrambled faces. The EEG signals were acquired with a 128 channel ActiveTwo system and downsampled from 2048 Hz to 200 Hz and averaged across trials, more specifications on the setup can be found in the SPM manual [10]. For reference we show the results using Friston et al.'s multiple sparse priors model (MSP) [3] as

¹ <http://www.fil.ion.ucl.ac.uk/spm/data/mmfaces/>

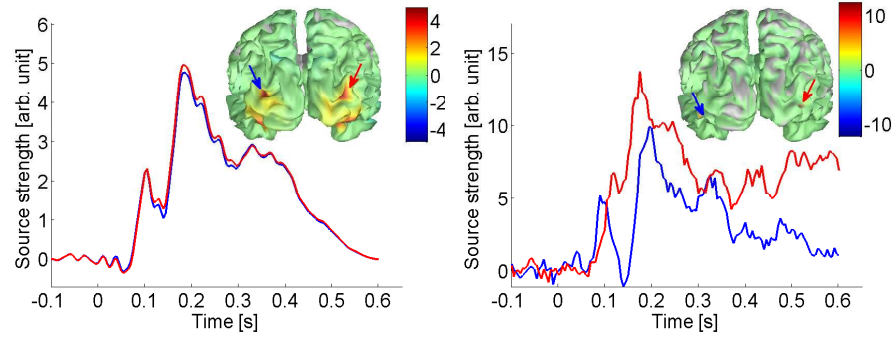


Fig. 4: The time evolution of the two strongest sources found in face response. In the inset of the cortex (posterior view) arrows point to the corresponding locations 180 ms post-stimuli. Left: Sources obtained using SPM's multiple sparse priors method. The color coding indicates that the solution obtained is rather dense. Right: Sources obtained using Zhang et al.'s T-MSBL method. As in MSP, estimate is rather dense.

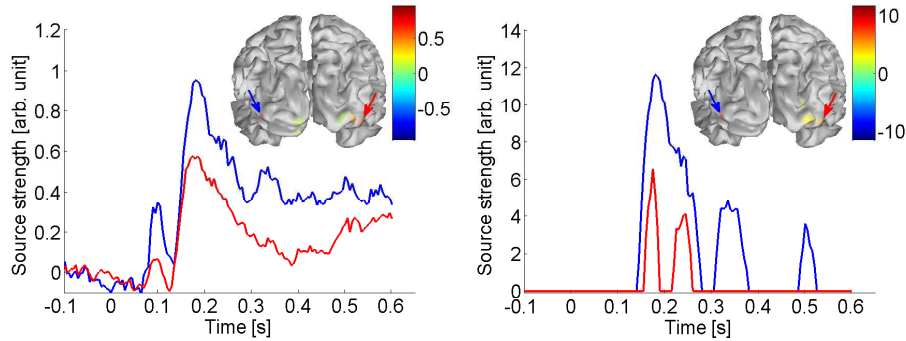


Fig. 5: The time evolution of the two strongest sources found in face response. In the inset of the cortex (posterior view) arrows point to the corresponding locations 180 ms post-stimuli. Left: Solution obtained using the temporally expanded VG (teVG) scheme. The color coding indicates more sparse solution than that obtained by MSP and T-MSBL. Right: Estimate by MarkoVG. As teVG, MarkoVG is also spatially very sparse. The time courses are, however, now sparse, thus the difference between face and scrambled face processing has been localized to focused shorter time intervals.

adopted in the SPM software and the result of Zhang et al.'s multiple measurements vector sparse Bayesian learning model, T-MSBL [4] in Fig. 4 left and right panels, respectively. These solutions should be compared to two versions of the VG: Time expanded VG (teVG) [9] and MarkoVG, both shown in Fig. 5. The MSP and T-MSBL estimates of the sources responsible for the *difference* between face and scrambled faces are very smooth in time, in fact resembling standard ERPs. This is also the case for the window-wise constant support model teVG, c.f., Fig. 5 (left), while the more flexible support recovery method MarkoVG finds a smaller number of active sites for the difference signal in Fig. 5 (right).

Also we note that teVG and MarkoVG in general find spatially sparser solutions, viz., the more extended gray areas in Fig. 4.

3 Conclusion

We have proposed MarkoVG assigning a Markov prior for promotion of temporally smooth sources in the Variational Garrote. We derived the modified variational update rules and identified the role of the Markov prior parameters. We showed that MarkoVG converges fast, as VG also does, and we found that sources are reliably estimated both in terms of location and source strength mean square error. In a benchmark EEG data set we showed that MarkoVG produced more focused activation than multiple sparse priors and temporal sparse Bayesian learning, both of which are more similar to our earlier VG generalization, teVG, which assumes constant temporal support in specified windows.

Acknowledgement We thank the Lundbeck Foundation for support through the Center for Integrated Molecular Brain Imaging (CIMBI).

References

1. Gorodnitsky, I. F., Rao, B. D., *Sparse signal reconstruction from limited data using FOCUSS: a reweighted minimum norm algorithm*. IEEE Transactions on Signal Processing, 45(3), pp. 600-616 (1997).
2. Wipf, D. P., Rao, B. D., *An Empirical Bayesian Strategy for Solving the Simultaneous Sparse Approximation Problem*. IEEE Transactions on Signal Processing, 55(7), pp. 3704-3716 (2007).
3. Friston, K. J., Harrison, L., Daunizeau, J., Kiebel, S. J., Phillips, C., Trujillo-Barreto, N., Henson, R. N., et al.: *Multiple sparse priors for the M/EEG inverse problem*, NeuroImage 39(3), pp. 1104-1120 (2008).
4. Zhang, Z., Rao, B.D.: *Sparse Signal Recovery with Temporally Correlated Source Vectors Using Sparse Bayesian Learning*, IEEE Journal of Selected Topics in Signal Processing, Special Issue on Adaptive Sparse Representation of Data and Applications in Signal and Image Processing 5(5), pp. 912-926 (2011).
5. Stahlhut, C., Mørup, M., Winther, O., Hansen, L.K.: *Simultaneous EEG source and forward model reconstruction (sofomore) using a hierarchical bayesian approach*. Journal of Signal Processing Systems, 65(3), pp. 431-444 (2011).
6. Montoya-Martinez, J., Artes-Rodriguez, A., Hansen, L.K., Pontil, M. *Structured sparsity regularization approach to the EEG inverse problem*. In Proc. 3rd International Workshop on Cognitive Information Processing (CIP 2012), IEEE (2012).
7. Kappen, H.: *The Variational Garrote*. arXiv preprint arXiv:1109.0486. Published at <http://arxiv.org/abs/1109.0486> (2011).
8. Hansen, S.T., Stahlhut, C., Hansen, L.K.: *Sparse Source EEG Imaging with the Variational Garrote*, 3rd International Workshop on Pattern Recognition in Neuroimaging (PRNI), pp. 106-109 (2013).
9. Hansen, S.T., Stahlhut, C., Hansen, L.K.: *Expansion of the Variational Garrote to a Multiple Measurement Vectors Model*, Scandinavian Conference on AI 2013 (2013).
10. Ashburner, J., Chen, C.-C., Moran, R., Henson, R. N., Glauche, V., Phillips, C.: *SPM8 Manual*, Available at <http://www.fil.ion.ucl.ac.uk/spm/doc/> (2012).